Learning to Rank for Geographic Information Retrieval

Bruno Martins and Pável Calado

GREASE-II Project Workshop

To be presented at the 6th Workshop on Geographic Information Retrieval

Geographic Information Retrieval

- **Given query *q* and set of docs $d_1, \dots, d_n$**
  - Documents seen a set of terms $d_i=(td_1,\dots,td_n)$ and a geo. scope *gd*
  - Query also seen a set of terms $q=(tq_1,\dots,tq_n)$ and a geo. scope *gq*
  - Find documents d_i relevant to *q*, accounting with terms and scopes
  - Typically expressed as a ranking on $d_1,\dots d_n$
  - GIR uses fixed similarity measures $sim(q,d_i)$ to rank results

- **Very large number of features available for GIR**
  - *e.g., textual similarity, geographic distance and containment, ...*
- **Hard to tune weighting coefficients by hand**
- ***Why not use Machine Learning?***

Learning to rank for GIR

- Queries + lists of docs with relevance assessments
  - Embedded in high dimensional feature space
    - Features are textual and geographical estimates of relevance
  - Labeled with relevance scores (e.g., binary assessments)
- Ranking as a supervised learning problem
  - Learn a similarity function between queries and docs
  - Maximize results for a given performance measure (e.g., MAP)

- We are given ($lst\text{-}ft_{q,d}$, $rel_{q,d}$) for q = {1..N}, d = {1 .. N_q}
- Where each element on $lst\text{-}ft_{q,d} = (x_{q,1}, \ldots x_{q,Nq})$
- And each feature vector $x_{q,n} = \langle ft_1, \ldots, ft_i, fg_1, \ldots fg_j \rangle$

Overview

- Motivation
- ***Related work***
  - *Geographic information retrieval*
  - *Learning to rank for information retrieval*
- The considered ranking features
- The SVM^{map} learning to rank method
- Experimental validation
- Conclusions and future work

Geographic Information Retrieval

- **Handling placenames over textual documents**
  - Leidner (2007) surveyed different disambiguation methods
  - Martins et al. (2010) proposed a machine learning method
  - ***Yahoo! Placemaker Web service***
- **Assigning documents to geographic scopes**
  - Anastácio et al. (2009) compared different methods
  - ***Yahoo! Placemaker Web service***
- **Retrieving documents according to geo. constraints**
  - Different methods experimented in the context of GeoCLEF
  - Frontiera et al. (2008) tested different geo. similarity methods

Learning to Rank

- Many different methods proposed in the literature
  - Pairwise methods : RankingSVM, RankBoost, ...
    - Reduce ranking to classification on document pairs
  - Listwise methods : AdaRank, **SVM^{map}**, ...
    - Directly operate on ranked lists to optimize an IR evaluation metric
  - Tutorial at the 2008 WWW conference by Tie-Yan Liu
  - Recent survey by Tie-Yan Liu (2009)

- Most works use the LETOR benchmark datasets
  - Documents, topics and relevance assessments from previous TREC experiments (OSHUMED and GOV2)
  - ***Data from GeoCLEF can be used to build a similar resource***

Overview

- Motivation
- Related work
- ***The considered ranking features***
  - *Textual similarity features*
  - *Geographic similarity features*
- The SVM^{map} learning to rank method
- Experimental validation
- Conclusions and future work

Textual Similarity Features

- Two *document* streams, i.e. *title* and *title+body*
- Title for each topic corresponds to the *query*

- **Standard text retrieval features**
  - Term Frequency and Inverse Document Frequency
  - Document length
  - Cosine measure with TF-IDF scores
  - BM25 score

Geographic Similarity Features

- **Area for the geographic scope of the query**
- **Area for the geographic scope of the document**

- **Measure of taxonomic distance by Jones et al. (2001)**

- **Measures of geospatial overlap between scopes**
  - Normalized overlap metrics surveyed by Frontiera et al. (2008)

- **Measures of geospatial distance between scopes**
  - Normalized distance by Martins et al. (2007)

Geo-Textual similarity

- Combined metrics (*ranking models*) corresponding to linear combinations of textual and geo. features

- Yu and Cai (2007) proposed a dynamic combination approach for GIR, where weight given to each component depends on the query specificity

- **BM25 + normalized distance (equal weights)**
- **BM25 + degree of overlap (equal weights)**
- **BM25 + degree of overlap (proportional weights)**

Overview

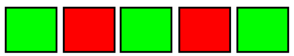
- Motivation
- Related work
- The considered ranking features
- ***The SVM^{map} learning to rank method***
- Experimental validation
- Conclusions and future work

Overview on SVM ^{map}

- Use Support Vector Machines for rank optimization

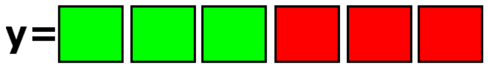
- Software available at http://svmrank.yisongyue.com

- Incorporate MAP in list-wise constraints (Yue et al., 2007)
  - Score w.r.t. true labeling should be higher than that w.r.t. incorrect labeling
  - Average precision is the average of the precision at the rank locations of each relevant doc.
  - Mean Average Precision (MAP) is the mean the Average Precisions for a group of queries

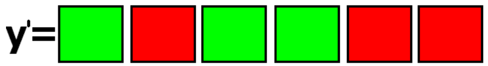
- Ex:  has average precision $\frac{1}{3}\cdot\left(\frac{1}{1}+\frac{2}{3}+\frac{3}{5}\right)\approx 0.76$

Challenges with SVM^{map}

- For optimizing Average Precision, the true labeling is a ranking where the relevant documents are all ranked in the front



- An incorrect labeling would be any other ranking



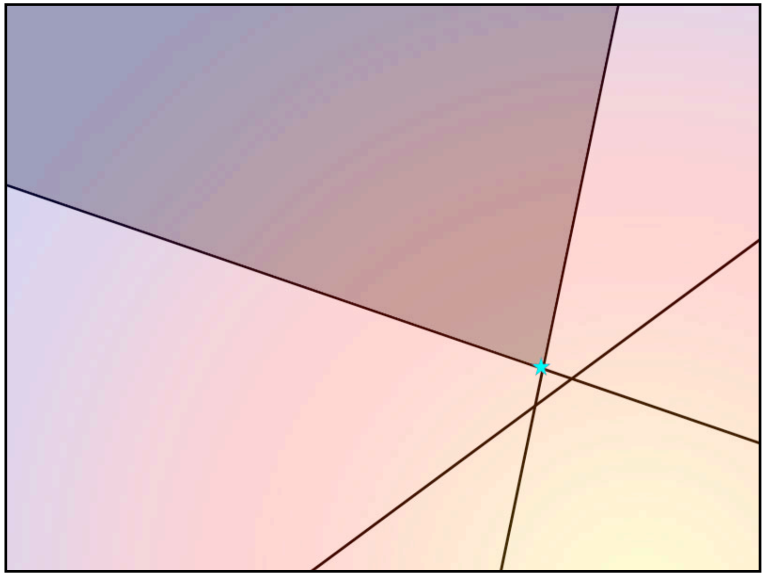
- Exponential number of rankings, thus an exponential number of constraints!

- Structural SVMs – efficient algorithm for tackling this issue, finding a small subset of important constraints.

Structural SVM training

[Tsochantaridis et al., 2007]

Structural SVM training proceeds incrementally by starting with a working set of constraints, and adding in the most violated constraint at each iteration



Original SVM Problem

- Exponential constraints
- Most are dominated by a small set of “important” constraints

Structural SVM Approach

- Repeatedly finds the next most violated constraint…
- …until a set of constraints which is a good approximation is found

Overview

- Motivation
- Related work
- The considered ranking features
- The SVM^{map} learning to rank method
- ***Experimental validation***
  - *The validation protocol using GeoCLEF datasets*
  - *Obtained results*
- Conclusions and future work

Experimental protocol

- **Data from 4 editions of the GeoCLEF evaluation campaign**
  - Total of 25 search topics in each edition (100 topics)
    - Geographically challenging topics (different "issues" included)
  - Newswire document collections
    - English collection: LA-Times 1994 and Glasgow Herald
  - Binary relevance judgments collected through pooling
- ***Documents + topics processed with Yahoo Placemaker***

- **Feature vectors extracted from the dataset**
- **Feature vectors used in a cross-validation setting**
  - Topics from three editions for training, remaining edition for testing

The GeoCLEF datasets

	2005	2006	2007	2008
Number of English Topics	25	25	25	25
Avg. Terms per Topic Title	6.64	5.76	6.08	5.48
Topics with correct scope	21	13	12	16
Topics with incorrect scope	5	6	8	5
Topics with no scope	0	1	3	3
Topics with small scope	9	14	11	9
Topics with medium scope	3	4	4	3
Topics with large scope	13	6	7	10
Relevant Topic-Doc. Pairs	1,028	378	650	747
Judged Topic-Doc. Pairs	14,546	17,964	15,637	14,528
Considered Topic-Doc. Pairs	18,000	18,000	18,000	18,000
Best Mean Avg. Precision	0.3936	0.3034	0.2850	0.3037

The Obtained Results

Tests with 8 different ranking approaches

Approach	Mean Average Precision 2005	2006	2007	2008	**Avg.**
SVM^{map} text	0.2841	0.2052	0.1637	0.1767	0.2074
SVM^{map} geo	0.1215	0.0377	0.0767	0.0827	0.0790
SVM^{map} text+geo	**0.3116**	**0.2131**	**0.1859**	**0.2024**	**0.2282**
TF-IDF	0.2293	0.1500	0.1541	0.1595	0.1732
BM25	0.2834	0.2052	0.1629	0.1761	0.2069
(BM25 + distance) / 2	0.2849	0.2063	0.1637	0.1786	0.2083
(BM25 + overlap) / 2	0.2838	0.2055	0.1636	0.1762	0.2072
BM25 + ($w_t \times$ overlap)	0.2856	0.2064	0.1636	0.1773	0.2082

- Geographic similarity alone performs poorly when compared to text retrieval
- When considering text-only approaches, BM25 outperforms TF-IDF
- Linear combinations of text+geo. are comparable to text-only approaches
- Learning to rank with text+geo. achieved the best performance

Overview

- Motivation
- Related work
- The considered ranking features
- The SVM^{map} learning to rank method
- Experimental validation
- ***Conclusions and future work***
  - *Conclusions*
  - *Future work*

Conclusions

- Learning to rank offers a principled approach for combining many different features
- Outperforms linear combinations of features
- Geographic similarity features do help in Geographic IR
- Other systems from GeoCLEF still had better results
  - Different number of documents retrieved for each query
  - Pseudo-relevance feedback, stemming, stop-words, ...

Future Work

- **Slightly different GIR problem formulation(s)**
  - Consider multiple geographic scopes per doc
  - Use all disambiguated place references
  - Measure impact of disambiguation performance!
- **Experiment with other learning to rank methods**
- **Ranking models specific for different queries**
  - Cluster queries according to specificity/locality
  - Classify queries as “local” versus “global”
- **Not all features are equally effective**
  - Feature selection methods specific for ranking

Thanks for your attention

bruno.g.martins@ist.utl.pt

SVM classifier adapted to ranking

- Input: **x** (candidate set of documents)
- Target: **y** (subset of **x** of size K)
- SVM objective function: $$\frac{1}{2}w^2 + \frac{C}{N}\sum_i \xi_i$$
- Constraints for each incorrect labeling **y**'.

$$\boxed{\forall \mathbf{y}' \neq \mathbf{y}: \quad w^T\Psi(\mathbf{x},\mathbf{y}) \geq w^T\Psi(\mathbf{x},\mathbf{y}') + \Delta(\mathbf{y}') - \xi}$$

- **Score** of best **y** at least as large as incorrect **y**' plus loss
- Weighted subtopic loss (later)